



Australian Bureau of Statistics

1352.0.55.019 - Research Paper: Modelling Detailed Business Operating Expenses from ABS Economic Collections (Methodology Advisory Committee), Nov 1998

Latest ISSUE Released at 11:30 AM (CANBERRA TIME) 05/11/1998 First Issue

Summary

Main Features

BACKGROUND MATERIAL

The background material for this session comprises excerpts from a working paper of the same title, which is currently being written.

Chapter 1 contains an overview of the working paper, and is included to give some perspective on the current state of this research.

Chapter 2 (incomplete) contains some relevant background information on ABS economic collections.

Chapter 3 contains the core of the material which is to be presented in the session. It introduces the motivation for using model-based estimation techniques, develops the theory underlying a number of simple models designed to capture the essential elements of the survey data, and provides a strategy for selecting the best model from the alternatives.

Some material on the issues of model prediction and the extension of the models to provide economy-wide estimates of detailed operating expenses will be presented in the session, but are not specifically addressed in the background papers.

AUTHORS NOTE

The approach to model-based estimation presented in this paper evolved during the course of an empirical review into a range of modelling approaches proposed by other ABS researchers. The models are essentially exploratory in nature, and are not put forward with the intention of supplanting the efforts of those who have the more onerous task of actually producing the official estimates. (Indeed it is unlikely that the more complex models could be made operational within a production environment.) The primary objectives of this research are to gain greater insight into the characteristics of the data, and to provide guidance to those responsible for collecting, processing and interpreting the data. I would strongly welcome any suggestions from the Committee which may assist in re-directing this research to better meet these objectives.

ISSUES SUBMITTED FOR CONSIDERATION BY THE COMMITTEE

A number of issues which have been raised by this research are presented below, under

three broad category headings. Each issue is accompanied by a number of supplementary questions and discussion points.

A. Policy issues

A.1 What are the potential advantages and disadvantages of model-based estimation (especially the use of auxiliary information) for ABS respondents and clients? How should ABS assess the net impact and determine whether to proceed?

- Will model-based estimation always result in "superior" estimates?
- What if the auxiliary data are of poor quality?
- Should more auxiliary data items be collected?
- What if the size of the survey is reduced?
- If "size" is shown to be a significant factor in determining the allocation of expenditures (as economic theory would suggest), how should the ABS deal with the representation of small, medium and large businesses in its collections?
- By "pooling" data collected in successive surveys, it may be possible to form better estimates of expenditure patterns, and possibly reduce sample sizes. What are the problems associated with this approach ? [eg. price effects, overlap of respondents, weighting.]
- Does model-based estimation compromise the usefulness of the estimates for subsequent economic analysis? For example, is it possible that the estimation procedure will "build-in" or strengthen relationships between some economic variables, while concurrently destroying other linkages?

B. Technical Issues - General

B.1 Are Committee members aware of any precedents and/or alternatives to the treatment of "missing" data put forward in this research ?

- The proposed modification of the multinomial model to accommodate "missing" data is a conceptually simple idea (although somewhat more complex to implement). It is likely that similar models have been fitted by other researchers $\frac{3}{4}$ although none have so far come to light.
- Is there is an easier way of accounting for "missing" data ?

B.2 The ABS is required to produce high quality, objective statistics. The use of model-based estimation methods has the potential to impose subjective assumptions upon the data. Conversely, the quality of the estimates may be queried if the ABS uses modelling techniques which appear to contradict established economic theory. Is it possible to meet both criteria ?

- The multinomial model (and variants) imply rather strong assumptions about the nature of expenditure patterns. Is this a problem ? Are there ways of minimising the problem (eg. grouping) ? Are there simple modifications which can be made to the models ?
- What diagnostic tests might be employed to check the distributional assumptions of the models ?
- Can the use of auxiliary variables be made atheoretic ?

B.3 How can the problem of identifying "missing" data be addressed ?

- The main economic surveys currently collect aggregate "other operating

expenses" as a single item. Respondents are not required to indicate which of the 25 component items they have included in their total.

- The accuracy of the prediction process could be improved considerably if respondents were required to indicate which expenditure items they have included in "other operating expenses".
- Alternatively, the catch-all category "other operating expenses" could be split, so that additional broad categories of expenditure (eg. Taxes & charges, Repairs & maintenance) can be more satisfactorily identified.
- Some expenditure items are reported by only a small proportion of businesses. When present, these items sometimes account for a significant share of total expenditure. It might be preferable to collect data on such items separately (or not at all).
- There would appear to be advantages in defining "other operating expenditure" to include only those expense items which are incurred *regularly* by the *majority* of businesses.

C. Technical Issues - Specific

Comments on any of the following specific topics would be welcome. Please don't restrict your comments to the accompanying dot points.

C.1 "Probability of selection" weights

- "Probability of selection" weights are clearly required at the prediction stage to produce economy-wide estimates. They have also been used throughout the modelling stage to combine contributions to the log-likelihood function. The basic idea is that the model should fit more closely those respondents which represent the largest number of businesses.
- The weights do not necessarily indicate that the response of any individual business is more characteristic of the wider population. This criticism may apply particularly where "missing" data are involved. There is perhaps a case for fitting unweighted models.

C.2 Post-stratification

- Stratification variables are generally determined by reconciling client needs with considerations of sampling efficiency. Model-based estimation, however, cannot be applied at the usual level of detail of such sampling schemes $\frac{3}{4}$ which may result in only one or two respondents per cell. That is, the model-based estimation approach must assume some degree of homogeneity across sample strata.
- Post-stratification can be implemented by either fitting the model separately to defined subsets of respondents, or (equivalently) by employing categorical explanatory variables. Serious problems of interpretation arise when such categorical variables are combined with auxiliary variables which are not specifically restricted to strata.
- Where post-strata have an inherent ordering (eg. "small" and "large") care will be required to maintain the continuity of fitted functional relationships across strata boundaries.
- The results from modelling may provide useful feedback to survey designers on similarities and dissimilarities between strata and on the need to adjust sample sizes.

C.3 Logit transformation

- The logit transformation will fit a curvilinear relationship between expenditure shares and the explanatory variables. This relationship tends to become more linear and monotonic as the number of categories increases. Is the logit transformation likely to be flexible enough to capture the expected functional relationships ? Are there potential problems at the extremes of the data (or asymptotically) ?
- By fitting the model to hierarchical groups of expense categories, it may be possible to achieve greater functional flexibility.

C.4 Grouping categories

- Modelling expenditure shares within hierarchical groups can undoubtedly lead to substantial computational efficiencies. However, the process of grouping data imposes additional restrictions on the model.
- It is possible that clients may be specifically interested in particular subsets of the model, and it may be advantageous to be able to detach parts of the model.
- Grouping may provide a means of containing problems with infrequently reported data items.

C.5 Imputation

- Imputed data cannot be used for model-based estimation, as the assumptions underlying the imputation process will almost certainly contradict the assumptions of the model.
- How could the model itself be used for imputation ?

C.6 Non-response

- Is it sensible to model the rate of non-response (a component of "missing" data) if the respondents to the detailed survey are subjected to more intensive follow-up than those in the main survey ?
- Should predictions be adjusted for *probable* non-response ?

C.7 Standard errors on predictions

- This topic has not yet been explored. Any comments or suggestions would be welcomed.

About this Release

About this Release

The background material for this session comprises excerpts from a working paper of the same title, which is currently being written.

Chapter 1 contains an overview of the working paper, and is included to give some perspective on the current state of this research.

Chapter 2 (incomplete) contains some relevant background information on ABS economic collections.

Chapter 3 contains the core of the material which is to be presented in the session. It introduces the motivation for using model-based estimation techniques, develops the theory underlying a number of simple models designed to capture the essential elements of the survey data, and provides a strategy for selecting the best model from the alternatives.

Some material on the issues of model prediction and the extension of the models to provide economy-wide estimates of detailed operating expenses will be presented in the session, but are not specifically addressed in the background papers.